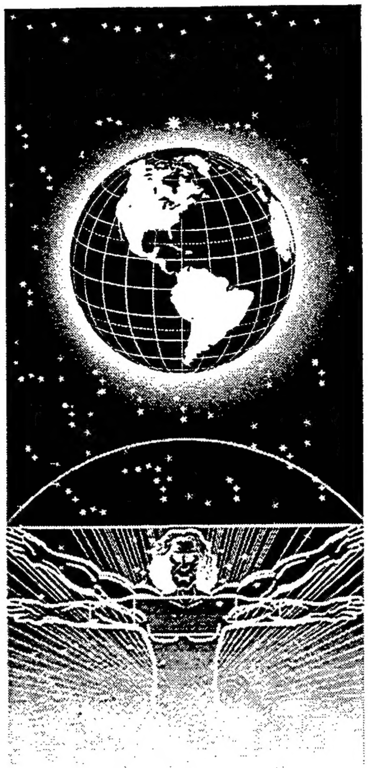
AFRL-HE-WP-SR-2001-0009



UNITED STATES AIR FORCE RESEARCH LABORATORY

The Intelligibility of Multiple Talkers Separated Spatially in Noise

Mark A. Ericson
Richard L. McKinley

May 2001

Final Report for the Period December 1987 to September 1993

20011016 168

Human Effectiveness Directorate
Crew System Interface Division
2610 Seventh Street
Wright-Patterson AFB OH 45433-7901

NOTICES

DISCLAIMER

This Special Report is published as received and has not been edited by the Air Force Research Laboratory, Human Effectiveness Directorate.

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-SR-2001-0009

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Instruction 40-402.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

MARIS M. VIKMANIS
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2001	Final - December 1987 to September 1993

4. TITLE AND SUBTITLE
The Intelligibility of Multiple Talkers Separated Spatially in Noise

5. FUNDING NUMBERS
PE - 62202F
PR - 7231
TA - 723139
WU - 72313901

6. AUTHOR(S)
Mark A. Ericson
Richard L. McKinley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Crew System Interface Division
Aural Displays and Bioacoustics Branch
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7901

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-HE-WP-SR-2001-0009

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Speech communications are seldom isolated auditory events in quiet environments. Frequently, the desired speech signal is confounded with other speech signals and noises. Real-world environments often degrade the intelligibility of the desired speech signal. In this book chapter (Binaural and Spatial Hearing in Real and Virtual Environments, Lawrence Erlbaum Associates, Mahwah NJ, publishers, 1997), the literature on the speech intelligibility of competing messages and the masking of speech is reviewed. The literature on the detection of speech is included to describe factors that can affect speech intelligibility. Following the review, several experiments are presented in which the effects of various conflicting signals on speech communications are measured. Virtual audio over headphones is used to investigate the effects of directional separation of talkers, the quantity and gender of talkers, the degree of masker interaural correlation, masking level, and selective attention. The results are discussed and compared with the previous literature.

14. SUBJECT TERMS
3-D auditory system, auditory localization, bioacoustics, bioengineering, frequency spectrum cues, synthesizer

15. NUMBER OF PAGES
26

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

This page intentionally left blank.

Chapter 32

The Intelligibility of Multiple Talkers Separated Spatially in Noise

Mark A. Ericson and Richard L. McKinley
Crew Systems Directorate, Armstrong Laboratory,
Wright-Patterson Air Force Base, Ohio

(Received December 1994; revised August 1995)

Speech communications are seldom isolated auditory events in quiet environments. Frequently, the desired speech signal is confounded with other speech signals and noises. Real-world environments often degrade the intelligibility of the desired speech signal. In this chapter, the literature on the speech intelligibility of competing messages and the masking of speech is reviewed. The literature on the detection of speech is included to describe factors that can affect speech intelligibility. Following the review, several experiments are presented in which the effects of various conflicting signals on speech communications are measured. Virtual audio over headphones is used to investigate the effects of directional separation of talkers, the quantity and gender of talkers, the degree of masker interaural correlation, masking level, and selective attention. The results are discussed and compared with the previous literature.

INTRODUCTION

Many real-life listening environments have a myriad of simultaneous competing auditory signals, much like in a cocktail party. One situation in which voice communication in poor listening environments is critical is in aircraft cockpits. In this situation, voice communication is sometimes difficult due to competing voice messages over the radio and/or intercom, low-fidelity speech signals, and high ambient noise levels. Many pilots monitor several radio channels simultaneously to navigate, to receive commands and clearances, and to maintain awareness of other nearby aircraft. Aircraft radios typically have limited bandwidth (approximately 3.5 kHz) and marginal speech-to-noise ratios (0 to 10 dB). Civilian commercial aircraft cockpit noise levels range from 85 to 100 dB SPL for most aircraft types, with some approaching the military aircraft noise levels of 95 to 115 dB SPL under normal operating (cruising) conditions. The safety of the pilot, the crew, the passengers, and people on the ground depend on the timely and accurate reception of voice information in an environment that is less than ideal.

A new technology has been developed that may have the capability of improving speech intelligibility, information transfer, and situational awareness in complex listening environments. Virtual or 3-D audio is a technology that can improve speech communication when there are competing messages. Virtual audio is realized by electronically simulating the natural binaural cues and creating the illusion of spatial auditory images. The effect can be created over headphones or loudspeakers, although only headphone presentations are considered in this chapter. Audio signals can be encoded with natural spatial cues to create the illusion of a sound appearing somewhere around the listener. The process causes the listener to perceive the sound to originate from a particular location outside his or her head. Without the spatial encoding process a listener hears diotic sounds as if they originate halfway between the two ears. Spatial or 3-D audio displays can be manipulated in azimuth, elevation, and distance. Virtual audio technology provides a flexible system for generating a virtual "cocktail party" presented via headphones. This development has enabled research on the cocktail-party effect and parameters affecting communication capability and performance. Previously, such research was cumbersome or impossible to accomplish with a physical sound system.

The focus of this chapter is to review the pertinent literature on speech intelligibility with competing messages, to quantify the effects of directional encoding on speech intelligibility, and to identify parameters affecting directional speech intelligibility. Directional speech intelligibility with multiple talkers is compared with diotic presentations of speech in quiet and in high-noise environments.

I. BACKGROUND

The following literature review is grouped into six general areas: (1) monaural aspects of speech intelligibility, (2) multichannel (left-eared and right-eared) presentations over headphones, (3) lateralized speech signals, (4) free-field talkers and maskers, (5) multipath interference, and (6) headphone presentations via manikins and synthesizers. Although some overlap does exist across these six categories, the grouping should enable discussion of several factors related to the cocktail-party effect. The review is intended to consolidate research findings of masking and binaural hearing with respect to their roles in understanding speech in real-world environments.

A. Monaural speech intelligibility

Before delving into the binaural aspects of listening to multiple talkers, a few comments should be made on the monaural aspects. A broad review on the masking of speech was written by Miller in 1947 and still is relevant today. The masking of speech by speech, noise, and tones was discussed. Monaural factors included intensity, spectrum, and temporal pattern of sound. Interruptions in the continuity of the masker's temporal pattern were found to decrease its effectiveness. Regardless of the type of sound, the spectra of the speech and noise were

the primary factors in the amount of masking. Based on this and other findings, the articulation index (Kryter, 1962) was developed to predict the percentage of speech intelligibility based only on the spectra of the speech and masker. Since this early work, other monaural and binaural effects on speech intelligibility have been investigated.

B. Multi-channel listening

Many everyday sounds interfere with speech communication. Cherry (1953) coined the term "cocktail party" to describe a typical situation in which speech can be understood despite several other sound sources. The interference may include other speech signals, music, mechanical noise, and transient auditory events. If a single microphone were immersed in the din of a cocktail party and recorded the sounds in the room, individual sources would be difficult to discern from one another when played back. If a manikin with a microphone in each ear were placed in the same location as the single microphone, then the individual talkers in the binaural representation would be more intelligible. Temporal and spectral information encoded by the manikin onto the speech signals would enable a listener to pay more attention to one auditory source of interest and suppress the others. Listeners in cocktail-party situations use monaural and binaural cues to attend to various audio signals (Miller, 1947; Cherry, 1953).

Cherry (1953) published his classic article on the improvements in speech intelligibility due to the separation of talkers into left and right channels. Several interesting observations were made. Contextual information facilitated the ability to follow a speech message that was heard among other messages. While following a particular message in one ear, unwanted speech or signals from the other ear could be more easily rejected than while following a string of words with no connected meaning. When asked to recall information about sounds heard in the ear opposite the speech message, only statistical information could be remembered. For example, the listener may recall the signal being speech, or noise, or a pure tone, but no other information. Cherry found that subjects could switch attention between talkers very quickly (up to seven times per second) without degrading understanding of the message. Although no spatial or directional properties were added to the speech signals, aspects of two-channel (two-eared) listening abilities in cocktail-party situations were examined.

Many other researchers began investigating other two-channel (two-eared) phenomena. Egan, Carterette, and Thwing (1954) found that equal intensities of speech in the two ears led to 50% intelligibility for a talker masked by himself. However, intelligibility values above and below 50% were found with two different talkers. Qualitative differences between the talkers would alter intelligibility levels due to pitch, dialects, and clarity of individual talkers.

Webster and Solomon (1955) varied the response complexity and applied information theory to quantify the benefits of two-eared listening. At low information-transfer rates large benefits for two-eared listening were found. However, at high transfer rates, the channel bandwidth limited the information going to each ear and little additional benefit was found for two-eared listening.

Broadbent and Ladefoged (1957) measured vowel recognition in the presence of other signals. Little additional advantage was found by separating the speech signals into two channels. They inferred from these results that the correlation in the binaural system was mostly effective with random or noncontextual signals. In other words, the peripheral signals of each ear were correlated with stored patterns in memory. The peripheral to central correlation was often more salient than left ear to right ear correlation. However, Broadbent and Ladefoged cautioned that generalizations of the experimental results to localizing speech in real environments would not necessarily be fruitful.

The two-channel listening experiments were important in evaluating factors involved in multitalker communications. In real-life situations, sounds aren't separated into two channels but overlap and blend across the two ears. Actual listening performance in everyday situations is degraded by the presence of sounds from different auditory events being simultaneously present in each ear.

C. Headphone presentation of lateralized speech signals

Lateralization experiments have demonstrated the relative effects of interaural level differences (ILDs) and interaural time differences (ITDs) on speech intelligibility. The perceived location of a lateralized sound is inside the head and along the interaural axis. Many researchers have investigated the effects of lateralization on speech intelligibility, beginning with Licklider (1948). In general, combined time and level differences were found to provide higher intelligibility level differences than either ITD or ILD alone. An ILD is usually described by a single value in decibels and is independent of frequency. Corbett (1986) spectrally filtered speech and noise signals into various frequency bands and presented the signals over headphones to a listener. Corbett found improvements in speech intelligibility using this technique. A variation on the ITD parameter was made by amplifying the time differences to greater than normal differences of about 800 μs. Kollmeier and Peissig (1990) found slight improvements in speech intelligibility using this technique. One advantage of lateralization experiments is the ability to individually control ITD and ILD parameters via headphone presentation. When sounds are generated away from a listener's head as in free-field conditions, the ILD and ITD cannot be individually controlled. The next section contains descriptions of speech intelligibility of multiple talkers in free-field environments.

D. Free-field listening

Free-field listening incorporates the monaural factor of the best ear signal-to-noise ratio (SNR) and the binaural factors of interaural time and interaural level differences. Compared to two-channel listening, absolute speech intelligibility performance in free-field listening is slightly degraded due to signal and noise being heard in both ears simultaneously. Relative performance within the free-field condition was found to be a function of spatial separation and frequency content.

Spieth, Curtis, and Webster (1954) found an increase in speech intelligibility with horizontal spatial separation and with shaping filters, for responding to one of two simultaneous competing messages. Spieth *et al.* also investigated the effects of context in which messages were presented. Clichés were used to couch speech information within meaningful fragments. Speech intelligibility was higher when left and right signals were switched between a cliché, so that the entire cliché was heard intact by the same ear, than when randomly switched within a cliché. One possible inference is that higher order cognitive processing was being incorporated when listening to meaningful phrases. Bregman and Campbell (1971) developed a theory of auditory streaming and auditory scene analysis related to the cocktail-party effect. Recently, Bregman (1990) expounded on the theory of auditory streaming.

Webster and Thompson (1954) investigated responding to both of two overlapping messages. On average, 20% of the time messages overlapped. Leading messages prevailed over lagging messages as measured by number of phrases correct. Total information transfer was increased if messages had low information content. These findings agreed with results of a later experiment by Webster and Solomon (1955).

In a series of five experiments, Dirks and Wilson (1969) measured speech intelligibility in the free field and via a Kunstkopf. Competing noises and competing messages were used to mask the speech signal. This article contained an excellent review of the literature at that time. Unfortunately, measurement of the cocktail-party effect has not progressed very much since then. Some recent work by Yost, Sheft, and Dye (1994) and Yost (1995) should provide some valuable, quantitative data to the literature.

E. The effects of multipath signals on speech detection

Adding reverberation to the competing message experiments as described in the previous free-field section provides another factor of the "cocktail-party" effect described by Cherry. The reflections from a listening environment have long been known to reduce the level of speech intelligibility (Haas, 1951). The precedence effect, as described by Haas, had a maximum echo suppression of about 10 dB at 15 ms after the first wavefront. The most degrading effect on speech intelligibility from a single reflection occurred after the maximum echo-suppression delay, at about 30 ms after the first wavefront. These experiments were conducted with a single talker's voice and only a single reflection. The inclusion of other reflections in reverberant environments successively degrades speech intelligibility by reducing the interaural correlation and the SNR.

Hirsh (1950) found that thresholds of speech intelligibility were raised when listeners were moved from an anechoic environment (61 dB) to a reverberant environment (66 dB). The latter condition is the worst case situation for a single talker in a highly reverberant environment. Head motion cues seemed to improve (lower) the threshold of speech intelligibility from 63 dB with a fixed head condition to 59 dB with the head motion condition. Multiple talkers tend to degrade speech communication performance even further than random noise, due to the similarity of speech signal spectra and modulations.

Tobias (1972) simulated an airborne "cocktail party" with speech presented over an array of three loudspeakers in a small aircraft. Competing messages were presented either over a single center loudspeaker or over two separate loudspeakers, either in phase or out of phase. Only a small benefit of 2 dB was measured for the out-of-phase separate loudspeakers compared to the single in-phase loudspeaker condition.

In general, speech discrimination is better with binaural hearing than with monaural hearing in reverberant environments. The binaural system serves to reduce the deleterious effects of reverberation on localization, as reported by Wallach, Newman, and Rosenzweig (1949). The "squelch effect" as observed by Koenig (1950) is a decrease in the perceived amount of reverberation when listening binaurally as compared to listening monaurally or diotically. Later, Koenig, Allen, and Berkley (1977) measured masking level differences of about 3 dB for both coherent and incoherent maskers in a reverberant environment. Mackeith and Coles (1971) measured the effects of reverberation on binaural and monaural speech discrimination. This work was mostly motivated by hearing aid research as to the benefit of two-eared versus one-eared listening. They found changes in the speech-to-noise ratio from 0 to 4 dB for the squelch effect depending on the locations of the speech and masker and degree of reverberation. As noted before, the binaural hearing system tends to provide its greatest advantage over the monaural system when listening conditions are degraded by competing sounds.

F. Headphone presentation of free-field directional cues

Schubert and Schultz (1962) conducted two experiments in which masked speech signals were more easily understood by listening binaurally than monaurally. In the first experiment, the speech was masked by broadband random noise. Three speech ranges were filtered and presented to the listener. Each of the three frequency ranges was presented at three interaural time differences. The interaural time difference conditions included homophasic, antiphasic, and a 0.5-ms delay. The low-frequency speech was observed to provide the highest intelligibility percent improvement over the homophasic condition. From these data, the auditory system was inferred to make use of longer periods (6–15 Hz modulation) in the speech waveform when masked by broadband random noise. Binaural fusion was conjectured to operate peripherally by extraction of the low-frequency modulation envelope of speech waveforms.

In the second experiment, speech of a single talker was masked by speech signals from various sets of talkers. The same interaural time difference conditions as in the first experiment were used. The antiphasic condition yielded slightly higher masking level differences than the delayed speech condition. Although, significant ($p < 0.01$) MLDs (masking level differences) were found for maskers of five simultaneous talkers, multiple random talkers, and the talker's own voice for both anti-phasic and delayed speech conditions, generally, the binaural system was less efficient at extracting speech information from speech-like maskers than from random noise maskers.

Schubert and Schultz (1962) hypothesized that one might expect the largest differences between monaural and binaural hearing for signal detection, next for localization, and least for identification of a signal. However, they noted that factors such as contextual information play a role in localization and identification due to pattern matching and fusing of harmonically coherent portions of the monaural spectrum. Therefore, data from signal detection experiments may not always coincide with speech intelligibility data.

Bronkhorst and Plomp (1988) used speech reception thresholds (SRTs) to measure effects of ITD, ILD, and a combination of these two factors for speech presented from a virtual location directly in front of the subject (0° azimuth) and noise presented at various virtual directions in azimuth. When the noise was synthesized with both ITDs and ILDs, thresholds were lower than when the noise contained only ILDs or only ITDs. The data were converted to binaural intelligibility level differences (BILDs) in decibels by subtracting the mean SRT for each condition from the mean SRT for 0° free-field noise. The sum of the BILDs for the ILD only (5.5 dB) and ITD only (4.6 dB) noise masking conditions was higher than that for the combined free-field (both ILDs and ITDs) condition (8.1 dB). An ILD effectively reduced the overall release from masking when it was introduced into the ITD-only noise masker. That is, a simple linear combination of ILD and ITD effects would have produced a 10.1-dB BILD, instead of the measured 8.1-dB BILD. Previous experiments in the free field (Plomp and Mimpen, 1981) agreed with the combined threshold data.

Bronkhorst and Plomp (1992) measured the effects of multiple speech-like maskers on SRTs for normal and hearing-impaired listeners. Interfering noise was modulated by speech waveform envelopes and spectrally matched to the long-term average spectrum of speech. On average, a 3-dB advantage was found for the binaural over the monaural mode. The monaural contribution was observed to be considerable when compared to the binaural advantage. However, the monaural and binaural contributions were strongly dependent on the number and azimuthal positions of the maskers.

Ricard and Meirs (1994) measured the intelligibility of speech from virtual directions in azimuth. Stimuli included synthetic speech and a 5-kHz white-noise masker without modulation. Thresholds for masking of speech were found by linear extrapolation to the 70% speech intelligibility level. On average, thresholds were reduced by 4–5 dB for speech presented at various directions in azimuth with the interference always straight ahead.

A model of the binaural advantages in speech intelligibility was developed by Zurek (1993). The model accounts for a single interfering sound source in azimuth located in an anechoic environment. Zurek's model distinguishes itself from other models by taking into account interactive effects found in binaural hearing. As data become available, other variables, such as multiple maskers, elevation angle, distance, and reverberation, will hopefully be included in future models. The current model and inclusion of other factors will help to predict speech intelligibility in real-life environments.

Overall the cocktail-party effect literature contains several consistent findings. Large advantages are found for binaural speech intelligibility when speech and

noise signals are presented from different directions in azimuth. The absolute contribution of the monaural cues is much larger than the absolute contribution of the binaural cues. The greatest monaural cue is the relative energies in the spectra of the speech and noise waveforms. Binaural hearing provides a relatively large advantage to speech intelligibility in low speech-to-noise ratio conditions. Contextual information tends to improve speech intelligibility but not speech detection. Binaural hearing in reverberant environments is more robust than monaural hearing due to the "squelch effect." Multiple speech-like maskers are more effective than broadband, random noise maskers due to low-frequency modulations of the speech waveform envelope.

II. METHODOLOGY

A. Facilities and equipment

Speech intelligibility performance was measured using either the coordinate response measure (CRM) (Moore, 1981) or the voice communications effectiveness test (VCET) (McKinley and Moore, 1989). Experiments were conducted in the voice communications research and evaluation system (VOCRES) (McKinley, 1979) and in the performance and communications research and technology (PACRAT) facility.

VOCRES includes a control room, a reverberation chamber, and 10 subject stations in the chamber. VOCRES's sound generation system is capable of producing up to an over all 130 dB (SPL) of broadband noise from 100 to 10 000 Hz. The chamber is 8000 ft^3 in volume with a reverberation time (RT_{60}) of 6 s at 500 Hz. Listening stations are equipped with individual AIC-25 intercommunication systems, compressed air regulators, alphanumeric displays, and response panels. Visual presentation of the sentences to the talker and collection of the listeners' responses are automated by an HP-9845 computer. Talkers wore an HGU-26/P helmet and an MBU-12/P oxygen mask, equipped with an M-169 microphone. The output from the microphone was transmitted by an AIC-25 intercommunication set to the input of Armstrong Laboratory's auditory localization cue synthesizer (ALCS) (McKinley, 1988).

ALCS units were installed in VOCRES to produce the azimuthal auditory display over headphones. The ALCS contained HRTFs from a KEMAR manikin measured at 1° spacings at 7 ft of radius. The ALCS operated in conjunction with a computer, head tracker, external audio source, and two-channel headphones. A Polhemus electromagnetic head tracker monitored the orientation of the listener's head, which was used to maintain a constant direction of the sound source with respect to the chamber. ALCS outputs were displayed over Bose AH-1A active noise-reducing headphones, configured for binaural operation.

PACRAT, like VOCRES, includes a control room, reverberation chamber, and 10 subject stations. PACRAT's sound system is capable of producing up to 137 dB of broadband noise from 16 to 10 000 Hz. The chamber is about 20 000 ft^3 in volume and with a reverberation time (RT_{60}) of 12 s at 250 Hz. Subject stations were equipped with the same equipment as in VOCRES plus three multifunction CRT displays to enter responses during the VCET task.

B. Subjects

A panel of 12 paid volunteer subjects, 6 male and 6 female, participated in the experiments. All subjects exhibited hearing sensitivities equal to or better than 15 dB hearing threshold level for audiometric frequencies from 125 to 8000 Hz. In addition, all subjects had normal middle ear function. All talkers were from the same geographic location and had the same Midwestern regional accent.

C. Procedure

Speech was either presented diotically, dichotically, or directionally over headphones. Diotic presentations were realized by mixing all signals together and presenting them equally to each earphone; these auditory images appeared to originate in the center of one's head. Dichotic displays of two talkers were made by passing one talker's voice to one earphone and the other talker's voice to the other earphone. Directional presentations of two-talker displays were achieved with one ALCS, and four-talker displays were achieved with two ALCS units. The speech signals were encoded for various directions around the listener in azimuth. Elevation angle was held constant at the horizontal plane. Distance cues were essentially absent. All signals were encoded with a constant gain term without multipath cues. Subjects were allowed to freely move their heads during testing; however, no gross amount of motion was visually observed during testing. The criterion measure was speech intelligibility as measured by either the CRM or the VCET.

The CRM is a nonstandardized test to measure the speech intelligibility of simultaneous talkers. Each test phrase contains a call sign, a color, and a number. Two call signs, "ringo" and "baron," were used. Talker call signs were randomized so that half (25/50) were for "ringo" and half were for "baron." Individual listeners were instructed to respond to either "baron" or "ringo" for each 50-phrase session. One of four possible colors included "red," "white," "blue," and "grey." Numbers ranged from "one" to "eight." A typical sentence embedded in a phrase might be "Ready Ringo, go to blue eight, now." If any one part of the response was wrong, then the entire phrase was scored as incorrect. There was no correction for guessing. Presentation of the test words was randomized. Talkers spoke equal numbers of the call signs "ringo" and "baron" within each session.

VCET was designed to measure the amount of information transfer in typical airborne communications. Words and phrases were based on typical radio communications aboard military aircraft. Phrases were generated by computer for each session from a 200-word vocabulary. Phrases were six words in length and formed meaningful, sensible thoughts. Information in bits for each transmitted phrase was predetermined. The average number of bits per received phrase was predetermined for each 44-phrase session. The information rate in bits per second was found after each session. Speech intelligibility scores were based on entire phrases being correct. Any portion of the phrase being incorrect made the scoring of that phrase incorrect. Talkers read the phrases once, without repetition. No correction for guessing was made.

In the first experiment, listening levels were predetermined and held constant through all sessions. The gain of the intercom was set to a constant level to provide the same speech-to-noise ratios across all presentation modes. To calibrate the gain, a 1-kHz, 1-V peak–peak sinusoid was input to the headphone amplifier. The sound pressure level under the earcup was adjusted to a fixed level (73 dB SPL) using a B&K 2131 spectrum analyzer, a flat plate coupler, a B&K 523 artificial ear, and a B&K 4145 pressure microphone. Sound pressure levels under each earcup were calibrated to within ±0.5 dB of each other.

In the other four experiments, listening levels were individually adjusted by the listeners to most comfortable levels. Each subject had a knob that adjusted the gain of the sidetone presented over a headset. A typical level was set 5–10 dB above the background noise. However, more experienced listeners tended to set their levels several decibels lower than the less experienced listeners.

III. EXPERIMENT 1: SPEECH INTELLIGIBILITY IN DIFFERENT DIRECTIONS

A. Method

Ten subjects from the 12-member panel were used. Either a pair of two males, two females, or a mixed male and female pair was chosen as talkers. A male and a female were assigned to each of the two (diotic and directional) listening conditions. All listeners participated in all conditions of the study.

Signals and maskers were set to predetermined levels. The speech-to-noise ratio was chosen to achieve speech intelligibility levels from near 100% to below 50%. Speech spectra from the three pairs of talkers and the noise spectra are shown in Figs. 1, 2, and 3. Peak speech energy is about 20 dB above the long term average speech spectra. The male and female speech spectra are the most

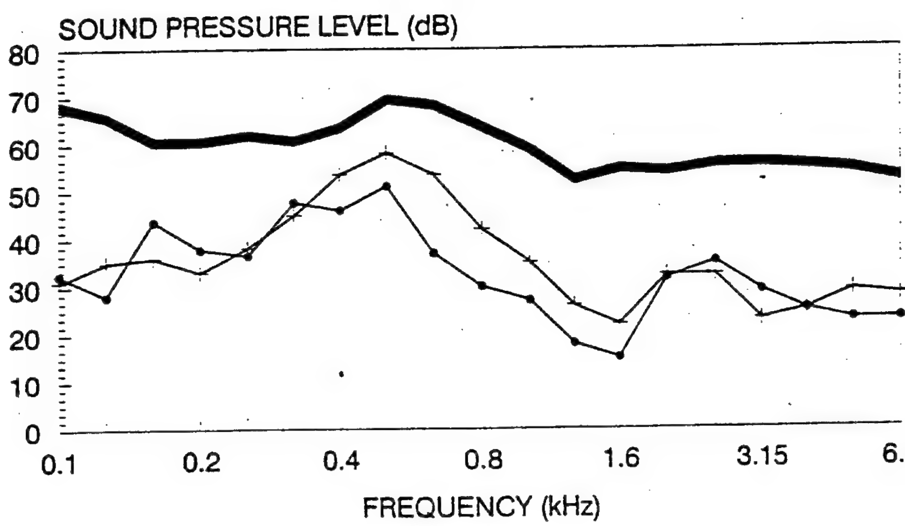


FIG. 1. Long-term average (32 s) of male speech (+), female speech (•), and 105 dB SPL noise (thick line) spectra.

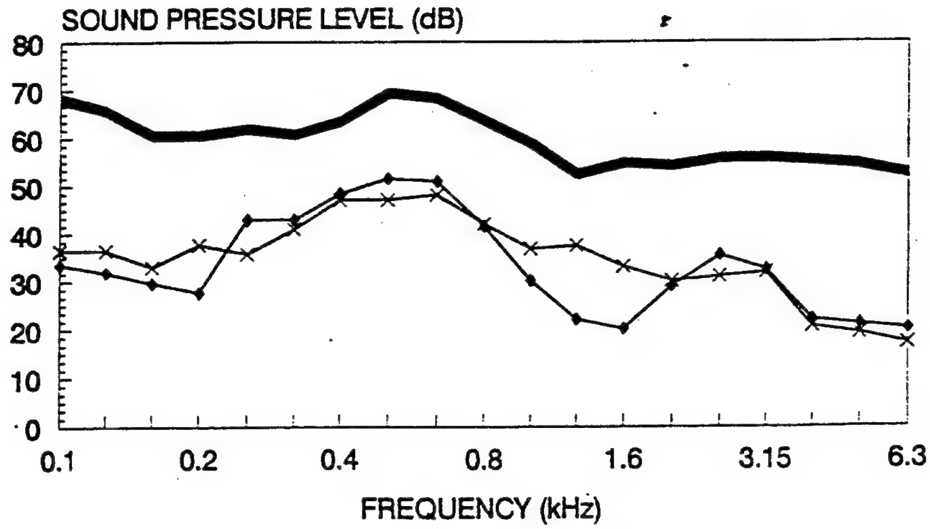


FIG. 2. Long-term average (32 s) of male speech (♦), male speech (x), and 105 dB SPL noise (thick line) spectra.

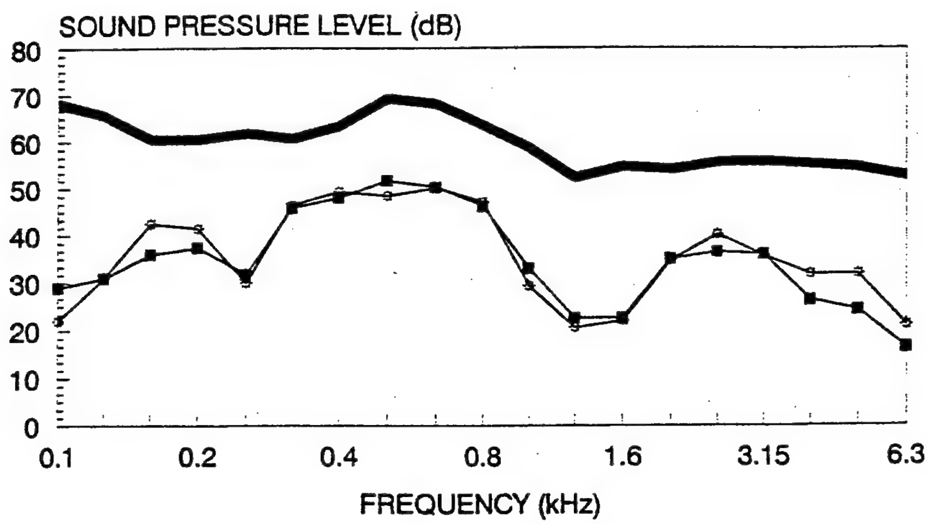


FIG. 3. Long-term average (32 s) of female speech (■), female speech (*), and 105 dB SPL noise (thick line) spectra.

dissimilar of the three pairs. The male speech spectra are very closely matched, except for between 1 and 2 kHz. The female speech spectra are the most closely matched of the three talker pairs.

The six noise levels included quiet (65), 85, 95, 105, 110, and 120 dB SPL. The spectrum and level of the noise under the headset were matched with a JBL one-third octave band graphic equalizer to the spectrum and level of the pink noise in the chamber. In this manner, the same signal-to-noise levels were realized for diotic and ambient maskers, although the interaural correlation of the maskers differed dramatically. Masking conditions of diotic, ambient, and a simultaneous combination of these two maskers were used to mask the talkers' voices. The diotic headphone masker had a correlation coefficient equal to 1.0. The ambient

masker in the chamber was estimated to have an average interaural correlation coefficient of about 0.3. Listeners perceived the reverberation to be very diffuse as also reported by Yanagawa, Anazawa, and Itow (1990). Lower frequencies tended to be more correlated than higher frequencies. The interaural correlation of the combined masker was between 0.3 and 1.0.

The CRM was used to measure speech intelligibility by the percentage of phrases correct. Three talker groups, six masking levels, three masking types, five separation angles, and two listening modes were repeated twice for a total of 1080 runs. Listener pairs ran in diotic and directional presentations for all experimental conditions to achieve a balanced experimental design.

B. Results

The interaural correlation of the masker had a measurable effect on speech intelligibility. Speech intelligibility was lowest with a diotic (high interaural correlation) masker. Speech intelligibility was highest with an ambient (low interaural correlation) masker. Speech intelligibility levels with combined maskers fell between the other two conditions. No interaction between the amount of masker interaural correlation and gender of the talker was observed.

In the quiet (65 dB) no masking condition, the effects of different talker genders were observed. Female voices tended to mask each other the most, producing the lowest intelligibility levels. Male voices masked each other less than female voices. Mixed-gender talkers masked each other the least. The relative effects of talker gender remained constant across all angles of separation.

Increasing angular separation improved intelligibility level differences between directional and diotic conditions. Zero degree nonseparation produced intelligibility levels the same as with diotic talker presentations. Small separations had a large effect on intelligibility. No additional benefit was found beyond 90° of separation. No interaction was observed between angular separation, talker gender, and masker correlation. Data for the 90° of separation condition from experiment 1 are graphed in Figs. 4, 5, and 6.

C. Discussion

In the first experiment, broadband noise maskers of three levels of interaural correlation were examined. The diotic masker, with high interaural correlation, was consistently observed to be the most effective masker of speech. Alternately, the ambient masker, with a relatively low interaural correlation, was consistently observed to mask speech the least. The masker correlation effect was seen within the three directional presentations and within the three diotic presentations. These differences were most prominent at the poor speech-to-noise ratios, that is, around the 50% intelligibility levels. Durlach (1964) measured the binaural masking level differences for different interaural correlations. The rank order of the intelligibility data agreed with the relative amount of masking for the various interaural correlations. Doll, Hanna, and Russotti (1992) measured improvements in masking thresholds when the background noise was uncorrelated with

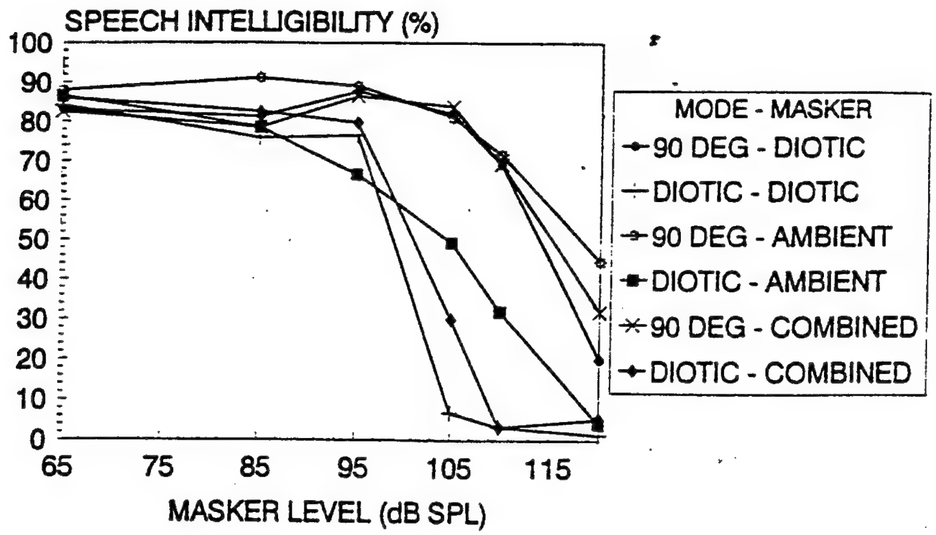


FIG. 4. Speech intelligibility of male and female speech versus masking noise level. Speech intelligibility was measured by the CRM at fixed presentation levels.

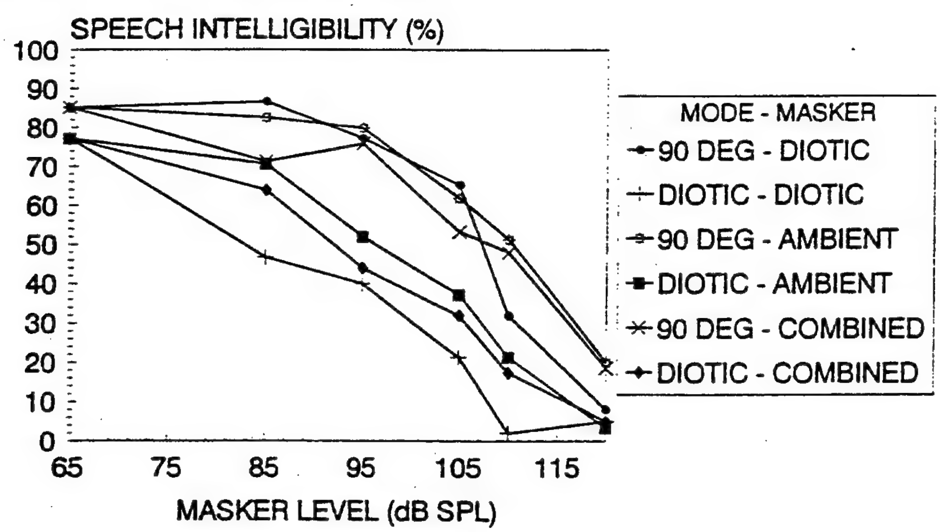


FIG. 5. Speech intelligibility of male and male speech versus masking noise level. Speech intelligibility was measured by the CRM at fixed presentation levels.

the signal and when angular separation increased. As are most factors observed in the cocktail-party effect, the degree of interaural correlation is a second-order effect after the primary factor, the speech-to-noise ratio at the better of the two ears.

Directional presentation of speech messages at 90° separation provides generally much higher intelligibility levels than with the diotic presentation. The binaural cues help to unmask the desired speech message from the interfering speech message and interfering noise. Within each presentation mode, the lowest intelligibility levels are measured with the diotic masker and the highest intelligibility levels are measured with the ambient masker. In Fig. 6, the intelligibility

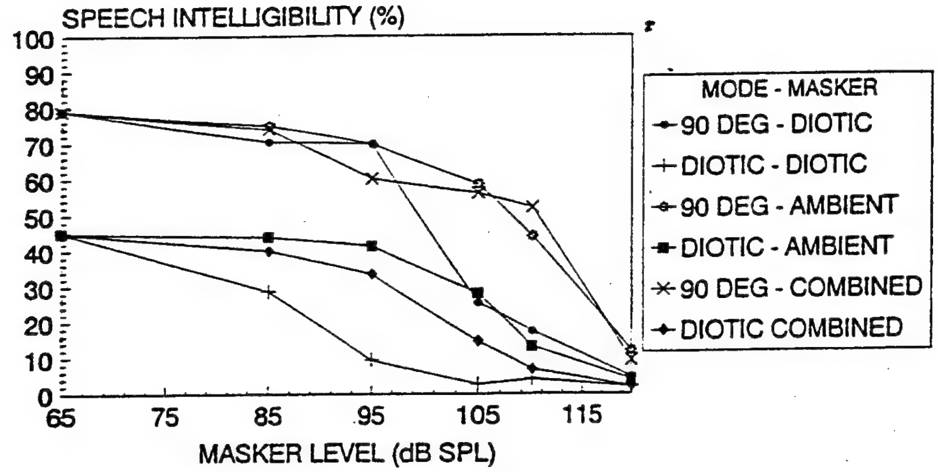


FIG. 6. Speech intelligibility of female and female speech versus masking noise level. Speech intelligibility was measured by the CRM at fixed presentation levels.

levels in the diotic presentation mode are lower than in Figs. 4 and 5 due to interference from the opposing female speech message. Presumably, the similarity of the female versus female speech spectra, similarity in the talkers' prosody, and similarity in quality cause more mutual interference than in the male versus male and male versus female speech conditions.

IV. EXPERIMENT 2: DIOTIC, DIRECTIONAL, AND DICHOTIC PRESENTATIONS OF SPEECH IN AMBIENT NOISE

A. Method

Speech intelligibility was measured for two competing messages using the CRM. In the dichotic test condition one message was presented to the left ear and the other to the right ear. In the directional test condition, talkers were directionally separated at one of five angles: 0°, 45°, 90°, 135°, or 180°. The control condition was the diotic presentation of both messages. The same subjects were used as in the first experiment.

Unlike experiment 1, the listener set talker voice amplifications to a most comfortable level. However, amplification of each talker channel was set to the same gain. No adjustments for different talker pairs were made. Talker pairs were chosen so that competing talkers spoke at similar loudness levels. One ambient, pink-noise masking level (105 dB SPL) and one quiet (65 dB SPL) level were used in VOCRES to provide speech-to-noise ratios representative of best and worst listening conditions.

A balanced repeated-measures design was employed. Three talker pairs, two masking levels, and seven listening conditions were repeated twice for a total of 84 runs. Listener pairs participated in diotic, directional, and dichotic presentations. Speech intelligibility levels for dichotic, directional, and diotic presentation modes were calculated for the two masking levels.

B. Results

Data for the second experiment are shown in Figs. 7 and 8 for 65 and 105 dB SPL ambient masking levels, respectively. Dichotic presentation of the competing messages was always more intelligible than the diotic presentation, and also more intelligible than 0° or 45° directional presentations. As was expected, speech intelligibility levels in the directional presentation condition at 0° were similar to levels in the diotic condition. However, a small angular separation of the messages (45°) greatly improved speech intelligibility. At 90° of separation, speech intelligibility levels were maximized and further separation did not yield higher intelligibility levels.

In quiet, female talkers tended to mask each other more than the male and mixed gender pairs. In ambient noise, the intelligibility of the dichotic presentation remained high (above 90%) compared to the levels in the diotic condition (62–84%). The speech intelligibility levels in the directional presentation condition at maximum separation approached those of the dichotic presentations.

C. Discussion

As shown again by the data of experiment 2, directional presentations of speech are more intelligible than diotic presentations, especially in low speech-to-noise environments. The same effects of angular separation and talker gender were observed in experiments 1 and 2. In practical situations, such directional presentations over headphones may improve speech communications when the signal is weak compared to the interfering noise, and the listener does not want to or cannot increase the signal level.

The dichotic (separate signals to the left and right ears) presentations provided higher levels of intelligibility than the small (45°) directional presentations. The left ear signal did not interfere with the right ear signal, or vice versa, in the

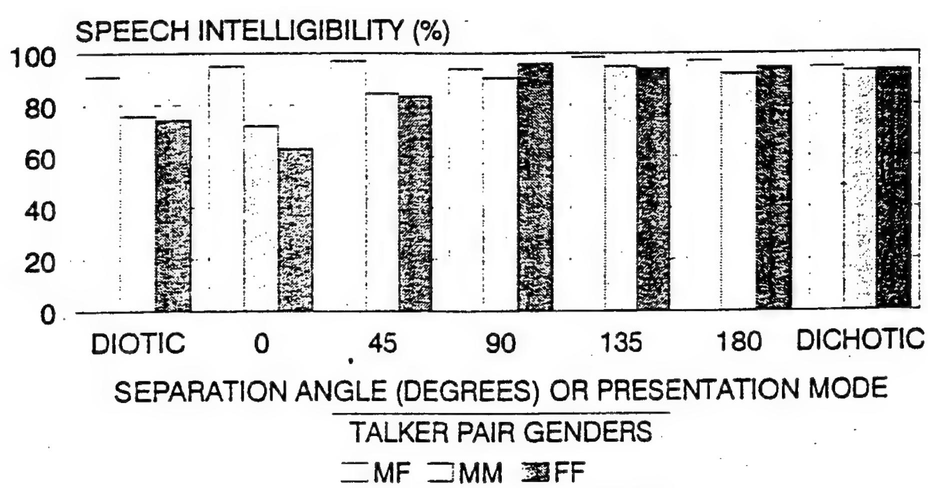


FIG. 7. Speech intelligibility for diotic, dichotic, and directional presentations of two talkers in quiet (65 dB SPL of ambient noise). Speech intelligibility was measured by the CRM with talker presentation levels set to most comfortable levels.

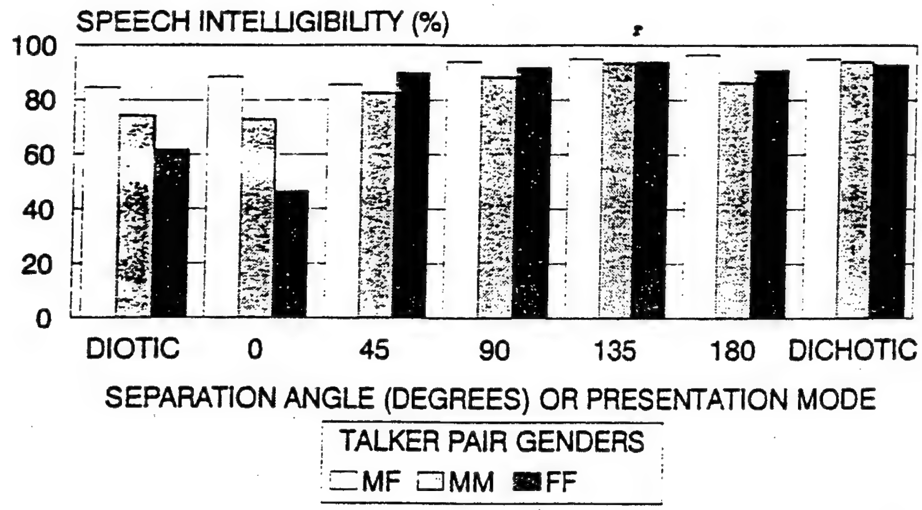


FIG. 8. Speech intelligibility for diotic, dichotic, and directional presentations of two talkers in 105 dB SPL of ambient pink noise. Speech intelligibility was measured by the CRM with talker presentation levels set to most comfortable levels.

dichotic presentation. However, the 45° directional presentation did contain these cross-talk signals, which produced lower intelligibility levels. The deleterious effects of the combined ITD and ILD cues were the same as measured by Bronkhorst and Plomp (1988) using speech reception thresholds. As conjectured much earlier by Cherry (1953), the ear closest to the sound source in free-field environments receives a greater signal than the ear away from the sound source. When there are several sound sources around a listener, these multiple signals reduce the speech-to-noise ratio at the ear closest to the desired talker. Thereby, the overall intelligibility level is reduced by the unwanted but necessary binaural signals. The potentially best benefit of directional over dichotic presentations should be found in displays that contain more than two talkers, because we only have two ears.

V. EXPERIMENT 3: INFORMATION TRANSFER AND SPEECH INTELLIGIBILITY

A. Method

A factorial experimental design for each talker group was chosen to determine which, if any, factors affected the information transfer rate and intelligibility level difference between directional and diotic presentations. Information transfer and speech intelligibility were measured together using the VCET. The same talkers and listeners were used as in the first two experiments. Two talker groups, two masking levels, two presentation modes, and two separation angles made 24 sessions in the study. Separation angles included no separation (0°) and 180° (±90°) of separation. The control condition was the diotic presentation of talkers. Noise levels included quiet (65 dB SPL) and 105 dB SPL of ambient pink noise.

B. Results

Directionally separated and diotic presentations of VCET yielded similar response times, 8.16 and 8.20 s, respectively. On average, a set of 44 phrases had 33 bits per phrase. In the directionally separated condition, 4.04 bits per second were communicated between talker and listener. Similarly, 3.86 and 4.02 bits per second were communicated in the diotic and 0° conditions, respectively.

Speech intelligibility percentages are graphed in Fig. 9 for quiet and in Fig. 10 for 105 dB SPL of noise. Speech intelligibility percentages were about the same with the VCET in experiment 3 as with the CRM in experiment 2 using the coordinate response measure. Speech intelligibility averaged about 85% with 180° angular separation in azimuth, and ranged from 55 to 85% with the diotic presentation. No practical difference was found between the 0° separation and the diotic condition.

C. Discussion

In experiment 3, response times for diotic and directional modes were the same, although intelligibility levels were higher for the directional presentations. Because subjects were not allowed to repeat messages, the average number of bits per second would actually be higher with the directional presentation compared to the diotic presentation condition if talkers repeated messages until all the information was transferred. An advantage for directional over diotic presentations may exist as a reduced number of times a talker has to communicate. Such an advantage would be important in time-critical situations.

Webster and Solomon (1955) observed that complex tasks tended to reduce the additional benefit of binaural presentations. Because the percent intelligibility levels were similar for both the CRM and VCET tasks, then listeners were

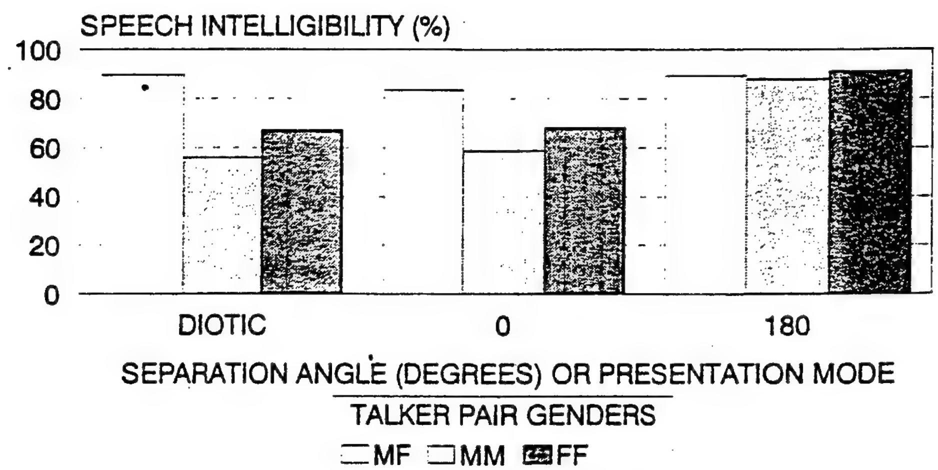


FIG. 9. Speech intelligibility for diotic, 0°, and 180° presentations of two talkers in quiet (65 dB SPL of ambient noise). Speech intelligibility was measured by the VCET with talker presentation levels set to most comfortable levels.

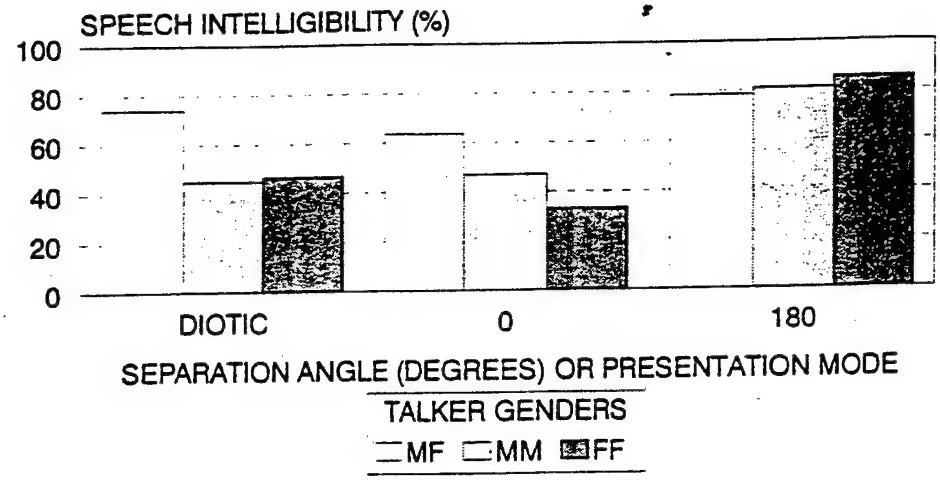


FIG. 10. Speech intelligibility for diotic, 0°, and 180° presentations of two talkers in 105 dB SPL of ambient pink noise. Speech intelligibility was measured by the VCET with talker presentation levels set to most comfortable levels.

probably not overtasked by the six word phrases in the VCET task. In other words, the binaural advantage was probably not limited by the width of the information channel.

VI. EXPERIMENT 4: FOUR COMPETING MESSAGES

A. Method

Speech intelligibility was measured for four competing messages using the coordinate response measure with two additional talkers. In the test condition each message was directionally separated by 0°, 30°, 60°, or 90°. For example, 30° separations placed talkers at 315°, 345°, 15°, and 45°. Likewise, 90° separations placed talkers at 45°, 135°, 225°, and 315°. The control condition was the diotic presentation of four messages. The third and fourth talkers functioned as distracters and went by the call signs "alpha" and "laker." The first and second talker call signs were randomized so that half (25/50) were for "ringo" and half were for "baron." Individual listeners were instructed to respond to either "baron" or "ringo" for each 50-phrase session. Twelve subjects participated in the experiment, eight as talkers and four as listeners.

Listeners set talker voice amplifications to a most comfortable level. However, amplification of all talker channels was balanced to equal gains for all talker groups. Talker groups were chosen so that competing talkers spoke at similar loudness levels. Listener performance was monitored to ensure that error rates were similar for each of the talkers. One ambient, pink-noise masking level (105 dB SPL) and one quiet (65 dB SPL) level were used in VOCRES to provide high and low speech-to-noise ratios.

The CRM was used to measure speech intelligibility in all experimental conditions. Three talker groups, three masking levels, four separation angles, and

two listening conditions were repeated twice for a total of 144 runs. Listener pairs participated in diotic and directional presentation modes to balance the experimental design.

B. Results

The same relative intelligibilities were found with four talkers as with two talkers. Overall levels were decreased due to the mutual interference of the competing three talkers. Only marginal intelligibility levels were achieved in the most optimal conditions (75% for MFMF at 90° in quiet). The addition of ambient pink noise greatly reduced speech intelligibility of the four talkers to barely intelligible levels. Data are plotted in Figs. 11 and 12.

C. Discussion

Data from experiment 4 showed little advantage for directional over diotic presentations of four simultaneous talkers. However, initial capture of the call sign may have been made easier by directional separation. The length of phrases made it more difficult to gain any advantage from initial capture. Yost *et al.* (1994) showed a benefit with single-word, multitalker experiments. Previous experiments showed benefit when less than seven talkers spoke unsynchronized phrases of different content (Bronkhorst and Plomp, 1992).

Less degradation in intelligibility would have been observed if the phrases had overlapped and had not been simultaneous. Four simultaneous talkers is an extremely difficult and unusual situation. There are not many situations in which one encounters monitoring four constant communications.

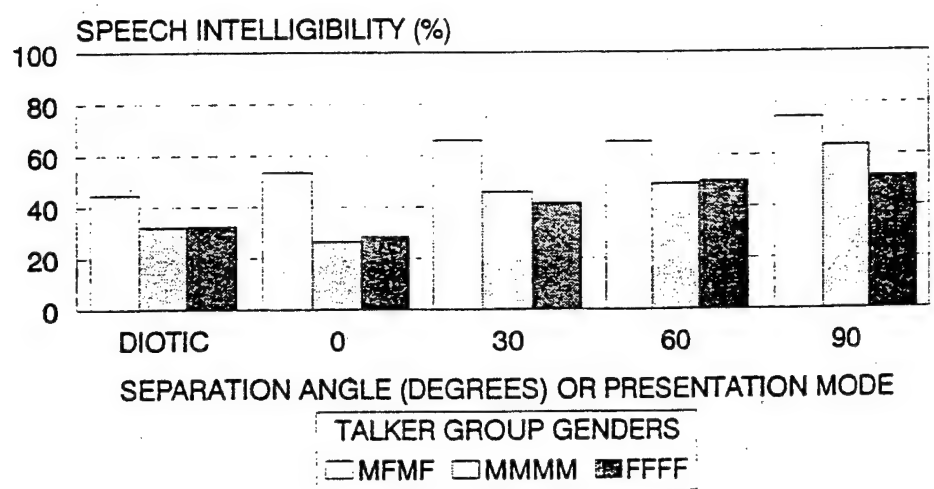


FIG. 11. Speech intelligibility for diotic and directional presentations of four talkers in quiet (65 dB SPL of ambient noise). Speech intelligibility was measured by the CRM with talker presentation levels set to most comfortable levels.

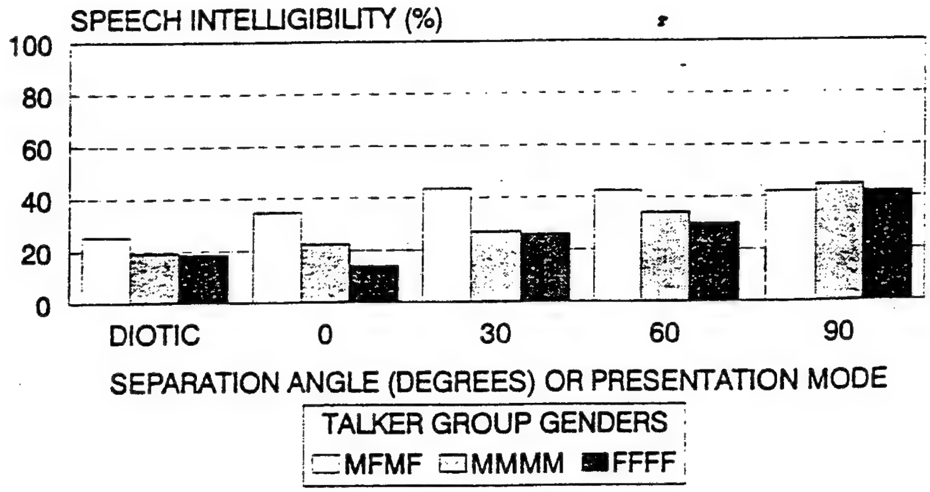


FIG. 12. Speech intelligibility for diotic and directional presentations of four talkers in 105 dB SPL of ambient pink noise. Speech intelligibility was measured by the CRM with talker presentation levels set to most comfortable levels.

VII. EXPERIMENT 5: SELECTIVE ATTENTION (TALKER LOCATION) AND SPEECH INTELLIGIBILITY

A. Method

The CRM was used to measure speech intelligibility for fixed versus random talker directions. In the fixed talker direction condition, listeners always heard the same talker's voice coming from the same direction. In the random talker direction, listeners did not know *a priori* from which direction a particular talker's voice would be heard. In this manner, the ability of the listeners to selectively attend to one direction could be measured. Two groups of talkers were used. Each group consisted of all male talkers or all female talkers. Talkers were chosen who spoke at similar loudness levels for each group. Four different phrases were used with four different call signs (ringo, baron, laker, and alpha). Directional separation angles included a control (0°) condition and a test (60° equal separation) condition. Talker voices were placed at 30°, 90°, 330°, and 270° in azimuth in the test condition. A total of 64 runs was made in quiet.

B. Results

No difference was found between the fixed and random directions. Angular separation improved speech intelligibility only 7% for male talkers and 5% for female talkers. Talker gender had no effect on speech intelligibility level differences. Overall speech intelligibility levels with the VCET were similar to previous four-talker conditions with the CRM. Data are plotted in Fig. 13.

C. Discussion

Selective attention to audio signals may be a fragile resource, one easily destroyed by multiple, simultaneous talkers. In other words, equal weighting may be

assigned to the start of every new message. Most data as described in the background section are on two talkers, as is often found in every day situations. Yost *et al.* (1994) observed a benefit of binaural displays with up to three talkers, but did not measure with four talkers. Bronkhorst and Plomp (1990) observed a benefit with up to six talkers, but the talkers spoke with pauses in an overlapping manner. The fifth experiment in this chapter was different from the others in that messages from four simultaneous talkers were heard by the listeners. Several simultaneous messages may overload the auditory system and prevent it from capturing the desired message from a particular direction.

VIII. GENERAL DISCUSSION

The cocktail-party effect cannot be measured by just one experiment. Unfortunately, one may infer from the name that there is a single cause, such as having two ears instead of one, that creates the effect. Hidden within the phenomenon are several factors that contribute to the overall ability to understand conversations in poor listening environments. Some of these include reflections in the listening environment, contextual information, prior knowledge of sounds, and quality of voices. The summation of all contributing factors may not add linearly, but interact, to provide an overall advantage greater than predicted.

A nonlinear relationship exists between speech intelligibility level and angular separation of talkers. The underlying reasons may be related to the way humans process binaural cues and the nonlinearity of the ILD and ITD functions in azimuth. The improvement seems to be most evident at low speech-to-noise ratios and in front of the listener where the ITD is at its steepest rate of change, about 10 μs per degree. Even a small talker separation (22°) centered in front of the

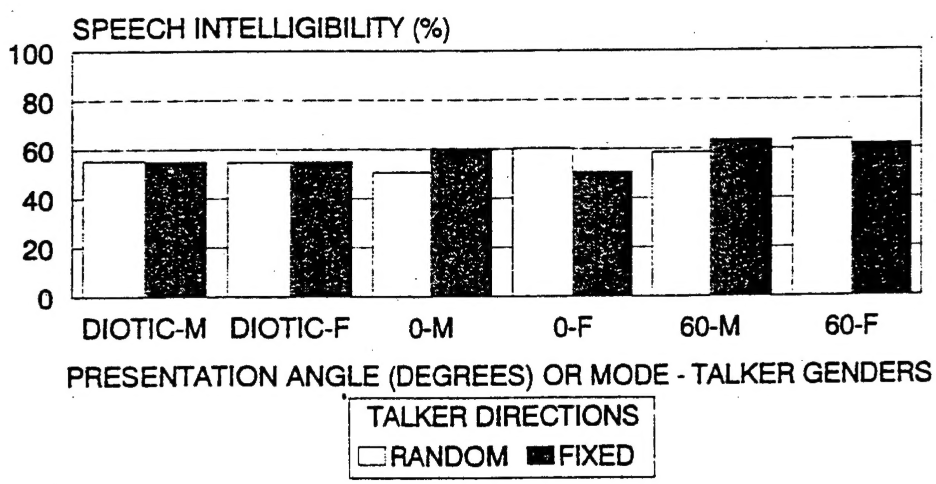


FIG. 13. Speech intelligibility for diotic, 0°, and 60° presentations of four talkers in quiet (65 dB SPL of ambient noise). Speech intelligibility was measured by the VCET with talker presentation levels set to most comfortable levels. In the 60° directional condition, talker messages were presented either from the same direction within each test session or from one of four random directions each time.

listener has a large effect on intelligibility. Talker separations centered at the side of the listener did not have as great an improvement for the same amount of separation.

Unwanted speech signals can act as maskers just as random noises do in the cocktail-party effect. Several attributes of speech signals affect the amount of disturbance on other desired sounds. The spectra of the speech signals are generally considered the most important factor in the mutual masking of speech. Pitch similarities across talkers play a role in the amount of masking. The female talker pairs in the experiments were observed to have very similar pitches and somewhat annoying timbre in their voices. In addition to their similar spectra, these factors reduced intelligibility, as is seen by comparing the data of the diotic presentations of the three talker pairs in quiet.

IX. SUMMARY AND CONCLUSIONS

The pertinent literature on speech intelligibility with competing messages was reviewed. The effects of directional encoding on speech intelligibility was measured and compared to speech intelligibility with diotic presentations. Several experiments were conducted in quiet, with maskers presented over headphones, and in high levels of reverberant noise.

Several parameters affecting directional speech intelligibility were identified. Overall the cocktail-party literature contains several findings consistent with the current experiments. The absolute contribution of the monaural cues is much larger than the absolute contribution of the binaural cues. The greatest monaural cue is the relative energies in the spectra of the speech and noise waveforms. Binaural hearing provides a relatively large advantage to speech intelligibility in low speech-to-noise ratio conditions compared to intelligibility in high speech-to-noise ratios. Speech-like maskers are more effective than broadband noise maskers due to low-frequency modulations of the speech waveform envelope. However, differences in speech waveforms, such as the amount of overlap and instantaneous differences, can cause other speech signals to be relatively poor maskers of a desired speech message. Large advantages are found for binaurally separated speech messages presented from different directions in azimuth. Perhaps the clearest benefit of having a binaural hearing system is to extract a single sound source direction from a cacophony of sounds, know where that sound is coming from, and better interpret meaning from that sound.

ACKNOWLEDGMENTS

The authors acknowledge the support and guidance of Drs. Charles W. Nixon and Thomas J. Moore during the development of this technology and the conduction of these experiments. The authors also acknowledge the efforts of Michael Ward, David Ovenshire, and Ronald Dallman during the data collection.

REFERENCES

Bregman, A. S., and Campbell, J. (1971). "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Exp. Psych. 89, 244–249.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Broadbent, D. E., and Ladefoged, P. (1957). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. 29, 708–710.

Bronkhorst, A. W. and Plomp, R. (1988). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. 83, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (1990). "A clinical test for the assessment of binaural speech perception in noise," Audiology 29, 275–285.

Bronkhorst, A. W., and Plomp, R. (1992). "Effect of multiple speech like maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. 92, 3132–3139.

Cherry, E. C. (1953). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. 25, 975–979.

Corbett, C. R. (1986). "Filtering competing messages to enhance mutual intelligibility," M.S. thesis, MIT, Cambridge, MA.

Dirks, D. D., and Wilson, R. H. (1969). "The effect of spatially separated sound sources on speech intelligibility," J. Speech Hear. Res. 12, 5–38.

Doll, T. J., Hanna, T. E., and Russotti, J. S. (1992). "Masking in three-dimensional displays," Hum. Factors 34, 255–265.

Durlach, N. I. (1964). "Note on the binaural masking level differences as a function of the interaural correlation of the masking noise," J. Acoust. Soc. Am. 36, 1613–1617.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (1954). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. 26, 774–782.

Haas, H. (1951). "Uber den Einfluss eines Einfachechos auf die Horsamkeit von Sprache [On the influence of a single echo on the intelligibility of speech]," Acustica 1, 49–58.

Hirsh, I. J. (1950). "The relation between localization and intelligibility," J. Acoust. Soc. Am. 22, 196–200.

Koenig, W. (1950). "Subjective effects in binaural hearing," J. Acoust. Soc. Am. 22, 61–62(L).

Koenig, A. H., Allen, J. B., and Berkley, D. A. (1977). "Determination of masking level differences in a reverberant environment," J. Acoust. Soc. Am. 61, 1374–1376.

Kollmeier, B., and Peissig, J. (1990). "Speech intelligibility enhancement by interaural magnification," Acta Otolaryngol. Suppl. 469, 215–223.

Kryter, K. D. (1962). "Methods for the calculation and use of the Articulation Index," J. Acoust. Soc. Am. 34, 1689–1697.

Licklider, J. C. R. (1948). "The influence of interaural phase relations upon the masking of speech by white noise," J. Acoust. Soc. Am. 20, 150–159.

MacKeith, N. W., and Coles, R. R. A. (1971). "Binaural advantages in hearing of speech," J. Laryngol. Otol. 85, 213–232.

McKinley, R. L. (1979). "Voice communications research and evaluation system," *Proceedings IEEE, National Aerospace and Electronics Conference, NAECON 79*, Vol. 1, p. 212.

McKinley, R. L. (1988). "Concept and design of an auditory localization cue synthesizer," master's thesis, AFIT/GE/ENG/88D-29, Air Force Institute of Technology, Wright-Patterson AFB, OH.

McKinley, R. L., and Moore, T. J. (1989). "An information theory based model and measure of speech communication effectiveness in jamming," *Proceedings of Speech Tech '89* (Media Dimensions, New York), pp. 101–105.

Miller, G. A. (1947). "The masking of speech," Psychol. Bull. 44, 105–129.

Moore, T. J. (1981). "Voice communication jamming research," Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings No. 311, Aural Communication in Aviation, CP311, 2-1-2-6.

Plomp, R., and Mimpen, A. M. (1981). "Effect of the orientation of the speaker's head and the azimuth of the noise source on speech reception thresholds for sentences," Acustica 48, 325–328.

Ricard, G. L., and Meirs, S. L. (1994). "Intelligibility and localization of speech from virtual directions," Hum. Factors 36, 120–128.

Schubert, E. D., and Schultz, M. C. (1962). "Some aspects of binaural signal selection," J. Acoust. Soc. Am. 34, 844–849.

Spieth, W., Curtis, J. F., and Webster, J. C. (1954). "Responding to one of two simultaneous messages," J. Acoust. Soc. Am. 26, 391–396.

Tobias, J. (1972). "Binaural processing of speech in light aircraft," FAA-AM-72-31, September.

Wallach, H., Newman, E. 3., and Rosenzweig, M.R. (1949). "The precedence effect in sound localization," Am. J. Psychol. 62 (3), 315–336.

Webster, J. C., and Thompson, P. O. (1954). "Responding to both of two overlapping messages," J. Acoust. Soc. Am. 26, 396–402.

Webster, J. C., and Solomon, L. N. (1955). "Effects of response complexity upon listening to competing messages," J. Acoust. Soc. Am. 27, 1199–1203.

Yanagawa, H., Anazawa, T., and Itow, T. (1990). "Interaural correlation coefficients and their relation to the perception of subjective diffuseness," Acustica 71, 230–232.

Yost, W. A., Sheft, S., and Dye, R. (1994). "Divided auditory attention with up to three sound sources: A cocktail party," J. Acoust. Soc. Am. 95, 2916.

Zurek, P. M. (1993). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.